Audio Quality Research

Outputs

- Technical publications and presentations demonstrating new research results.

- Objective estimates and subjective measurements of speech and audio quality.

- Algorithms and software for speech and audio coding and quality assessment.

Digital coding and transmission of speech and audio signals are enabling technologies behind many innovations in telecommunications and broadcasting including digital cellular telephone services, voice over Internet protocol (VoIP) services, and digital audio broadcasting systems. Speech signals can be coded and transmitted at rates as low as 4 kbit/s with good resulting quality. More general audio signals that include music and other sounds can be coded and transmitted with remarkably high fidelity at rates between 16 and 256 kbit/s per channel. In addition, coded speech and audio signals can be packetized for transmission, thus sharing radio spectrum or wired network bandwidth with other data streams and hence with other users.

In digital coding and transmission, one generally must trade off quality, bit-rate, delay, and complexity. In addition, the robustness of digital coding and transmission algorithms is critical in applications that use lossy channels. Important examples of lossy channels include those provided by wireless systems and those provided by the Internet. The ITS Audio Quality Research Program seeks to identify and develop new approaches that increase quality and robustness or lower the bit-rate, delay, or complexity of digital speech and audio coding and transmission. The ultimate result of such progress should be better sounding, more reliable, more efficient telecommunications and broadcasting services at lower costs.

In most digital speech and audio coding and transmission systems, a set of complex time-varying interactions among signal content, source coding, channel coding, and channel conditions make it difficult to define or measure speech or audio quality. The Audio Quality Research Program operates a subjective testing facility and runs controlled experiments to gather listeners' opinions of the speech or audio quality of various coding and transmission systems. The program has also developed and verified tools for the objective estimation of telephone bandwidth speech quality. Throughout FY 2003, subjective and objective audio quality testing was conducted to support Audio Quality Research efforts. In addition, customized objective speech quality estimation tools were developed to support other ITS efforts that are focused on the characterization of commercially available communications systems. Some of the laboratory equipment used to in the ITS Audio Quality Research Program is shown in Figure 1 below.

Two additional FY 2003 research efforts are briefly summarized here. In packetized speech transmission systems (the most prominent example is VoIP) transmission delay can vary significantly and rapidly, even within a single spoken phrase. This delay

Figure 1. Some of the laboratory equipment used to support the ITS Audio Quality Research Program (photograph by S. Wolf).

variation arises from the basic nature of packetized data networks and can be mitigated, but not eliminated, through buffering techniques. To understand the resulting speech quality, it is imperative that this continually changing delay be accurately tracked, and program staff worked to develop an algorithm to do this.

This algorithm must compare the input and output signals of the speech transmission system under test. But many systems of interest will distort speech waveforms, so a conventional waveform correlation solution often fails. The new algorithm uses speech envelopes that are generated by rectifying and low-pass filtering speech waveforms. High frequency information is lost in this stage, but significant robustness to waveform distortion is gained, and speech envelopes have proven very useful for determining coarse estimates of delay. To refine those coarse estimates, the next stage of the algorithm compares speech power spectral densities since these representations retain important properties of the speech signal, even when the waveforms suffer significant distortion. Figure 2 (above right) shows speech waveforms, speech envelopes, and speech power spectral densities for example input and output speech signal segments.

In a separate effort, program staff worked towards more robust speech coding through the method called multi-descriptive coding (MDC). In MDC an encoder forms multiple partial descriptions of a speech signal and these descriptions are sent over different physical or virtual channels. The MDC encoder does not know which of the channels are working and which of the channels have failed at any given time. On the other hand, the MDC decoder will know which of the channels have worked. If all descriptions arrive at the decoder intact, a higher-quality reconstruction of the speech is possible. If channel failures cause any of the descriptions to be lost, then a lower-quality reconstruction of the speech signal is still possible.
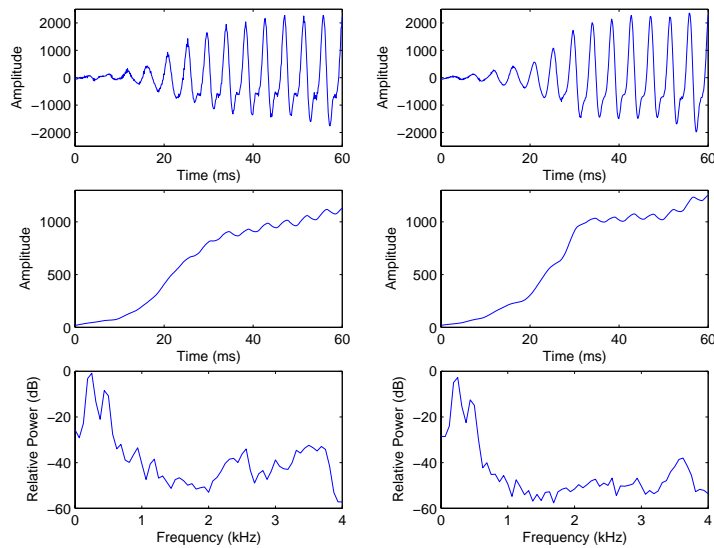

Figure 2. Speech waveforms, envelopes, and power spectral densities (top to bottom) are used to estimate the time-varying delay between a system input (on left) and output (on right).

Throughout FY 2003, the Audio Quality Research Program staff continued with selective upgrades to the ITS Audio-Visual Laboratories, including the introduction of a 5.1 channel digital audio system. The Audio Quality Research Program continued to transfer technology to industry, Government, and academia throughout FY 2003. Program staff prepared publications, delivered invited lectures and presentations, provided laboratory demonstrations, and completed peer reviews for journals and workshops. More detailed Program results are available at
http://www.its.bldrdoc.gov/home/programs/audio/audio.htm

Recent Publications

S.D. Voran, "Channel-optimized multiple-description scalar quantizers for audio coding," in *Proc. IEEE 10th Digital Signal Processing Workshop*, Pine Mountain, GA, Oct. 2002.

S.D. Voran, "Perception of temporal discontinuity impairments in coded speech – A proposal for objective estimators and some subjective test results," in *Proc. MESAQIN (Measurement of Speech and Audio Quality in Networks) Conference*, Prague, Czech Republic, May 2003.

For more information, contact:
Stephen D. Voran
(303) 497-3839
e-mail svoran@its.bldrdoc.gov